

# *R Text Mining Solution* (*R.Temis*) en el análisis avanzado de documentos de texto

---

**Prof. Dr. José Pino-Díaz**

Universidad de Málaga

Escuela de Ingenierías Industriales

Departamento de Economía y Administración de  
Empresas

Málaga, 21 de febrero de 2019



# CAQDAS

Acrónimo de **Computer-Aided Qualitative Data Analysis**, Software de Análisis Cualitativo de Datos Asistido por Computadora.

Méthode qualitative

- Logiciels
  - Réflexifs (CAQDAS)
  - Registres
  - Libres
- Formations

Christophe Lejeune

## Méthodes qualitatives informatisées

Conçu par [Christophe Lejeune](#)  
[This page in English](#)

De nombreux chercheurs en sciences humaines utilisent leur ordinateur comme un véritable laboratoire d'analyse de corpus de textes. Les progrès de la micro-informatique autorisent de tester des opérations calculatoires jadis confiées aux centres de calcul. La facilité et le moindre coût engendré permettent désormais d'envisager toutes les expérimentations concevables par l'esprit. Le coût résiduel concerne le temps que le chercheur est prêt à investir dans la conception, l'apprentissage ou l'usage de tels outils.

Ce site propose un panorama des logiciels disponibles dans le monde scientifique francophone. Cet exposé concerne les méthodes qualitatives : il s'agit des logiciels visant à **assister l'analyse de contenu, l'interprétation de récits de vie ou d'entretiens en sociologie qualitative**. J'ai délibérément favorisé la présentation d'outils :

- qualitatifs (au détriment des méthodes strictement quantitatives)
- francophones (parce que lesdites méthodes sont souvent sensibles à la nature linguistique du matériau. Dans la plupart des cas, les scientifiques francophones ont donc besoin d'outils adaptés à la langue (et la culture - !) des pays francophones.

Quelques exceptions échappent à ce principe ; il s'agit surtout de méthodes répandues de manière internationale (y compris en francophonie) qui auraient manqué au tableau. Leur diffusion m'a incité à les évoquer.

[Introduction](#)

- Alceste •AntConc
- Anvil •Atlas •Atlas
- Tétralogie •ATLAS.ti
- Calliope •Candide
- Caramel •Cassandra
- Catma •Civilité
- Coocs •Dedoose
- Discourse Network
- Analyzer •Dtm-Vic
- Elan •EthOS
- FreeQDA •General
- Inquirer •Glossanet
- Graphviz •Hamlet
- Hyperbase
- HyperRESEARCH
- Intex & Nooj
- IRaMuTeQ •Kwalitan
- Le graphe socio-
- technique •Lexico
- Leximappe
- LibreQDA •MaxQDA
- Modalisa •Morphix-
- NLP •Mosaïqs •Netvis
- Nodepad •NVivo
- Porphyry •Prospero
- Provalis •R •R.TeMiS
- Réseaux •RODA
- Sampler & LexiQuest
- Sato •Saturate
- Semato •Sonal
- Sphinx Lexica
- Squanto •T-Lab
- TamsAnalyser
- Textable •Textquest
- The Ethnograph
- Transana
- Transcriber •Trivium
- Tropes •TXM •Unitex
- Voyant Tools •Weft
- QDA

<http://www.squash.ulg.ac.be/logiciels/>

## ¿Qué es *Text Mining*?

La minería de textos es un área de la informática relacionada con el procesamiento del lenguaje natural, la minería de datos, el aprendizaje automático, la recuperación de información y la gestión del conocimiento.

La minería de textos se emplea para extraer información útil de datos textuales no estructurados a través de la identificación y exploración de patrones interesantes<sup>1</sup>. Forma parte del proceso KDD (*Knowledge Discovery Databases*).

La minería de textos permite ahorrar tiempo y dinero.

1 Bharati N Kharada en "Text Mining and its applications, Academia.edu; recuperado el 1 de febrero de 2019, disponible en [http://www.academia.edu/6888464/Text\\_Mining\\_and\\_Its\\_Applications](http://www.academia.edu/6888464/Text_Mining_and_Its_Applications), citando a Hotho A, Nürnberger A, Paaß G & AiS F (2005). A Brief Survey of Text Mining, recuperado el 1 de febrero de 2019, disponible en <https://pdfs.semanticscholar.org/9643/0cc91ed91fd2d4042fa6fcb7ecf4005d77a7.pdf>

# Herramientas de *Text Mining* y campos de aplicación

	information extraction	topic tracking	summarization	categorization	clustering	concept linkage	information visualization	question answering
<b>Medical:</b>								
FAQ's	x			x		x		x
Drug design	x				x	x		
New treatment		x				x		
<b>Business:</b>								
Competitive Analysis		x	x					
Media impact / analysis		x						
Current Awareness		x						
Intellectual property infringement	x	x			x			
Customer support for FAQ's	x			x	x			x
Social network detection							x	
Content personalization		x			x			
<b>Government:</b>								
Homeland security: detecting terrorist networks	x	x			x	x	x	
Law enforcement: crime detection / prevention	x	x			x	x	x	
<b>Education:</b>								
Research on a topic		x	x	x				
Citation analysis	x				x		x	
FAQ's	x			x	x			x

Table 3. Some examples of where text mining tools can be applied to the fields of medicine, business, government, and education.

1 Fan W, Wallace L, Rich S, Zhang Z (2005). "Tapping Into the Power of Text Mining"; recuperado el 1 de febrero de 2019, disponible en <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=64E863EEAA4C8B00E627E98E4B74707D?doi=10.1.1.81.1579&rep=rep1&type=pdf>

## *Text Mining* y sus aplicaciones<sup>1</sup>

1. Análisis de respuestas de encuestas abiertas.
2. Procesamiento automático de mensajes y correos electrónicos.
3. Análisis de reclamaciones de garantía o seguros, entrevistas de diagnóstico, etc.
4. Investigar competidores rastreando sus sitios web.
5. Mejora de la búsqueda web.
6. Análisis de patentes.
7. Clasificación de textos para agencias de noticias.
8. Bioinformática.
9. Filtrado Anti-Spam de Correos Electrónicos.
10. Minería de textos en bases de datos bibliográficas.

<sup>1</sup> Bharati N Kharada en "Text Mining and its applications, Academia.edu; recuperado el 1 de febrero de 2019, disponible en [http://www.academia.edu/6888464/Text\\_Mining\\_and\\_Its\\_Applications](http://www.academia.edu/6888464/Text_Mining_and_Its_Applications)

# Text Mining con R

## Módulo 1. Introducción: R software con licencia libre.



R es el software estadístico con licencia libre más extendido a nivel mundial tanto para docencia como para investigación.

Actualmente (08/02/2019) se encuentran disponibles más de **13.691 paquetes** desarrollados en R, que cubren multitud de campos científicos y tecnológicos.

Toda la información sobre R está disponible en la web <https://www.r-project.org/>

Existen interfaces gráficas que facilitan el uso de R: **RStudio**, **RKward** y **R Commander** se encuentran entre ellas.

---

### Recursos docentes del módulo



The R Project for Statistical Computing

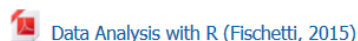
Sitio web del Proyecto R para Estadística Computacional ([The R Project for Statistical Computing](#)).



Videotutoriales de R y R Commander

---

### Recursos complementarios



Data Analysis with R (Fischetti, 2015)

**Data Analysis with R (Fischetti, 2015)**. Reciente monografía en inglés de Tony Fischetti sobre análisis de datos con R, editado por Packt Publishing (primera edición 22 de diciembre de 2015).



R for Data Science (Dan Toomey)

**R for Data Science** (Dan Toomey). This e-book is written for data analysts who have a firm grip over advanced data analysis techniques. Some basic knowledge of the R language and some data science topics is also required. This book assumes that you have access to an R environment and are comfortable with the statistics involved.



Statistique textuelle (Lebart y Salem, 1994)

**Statistique textuelle (Lebart y Salem, 1994)**. Manual francés clásico de estudio de la estadística lexical.

1 Modulo 1 del curso “Empleo de R *Text Mining Solution* en el análisis avanzado de documentos de texto”; cursos on line de la FGUMA (año académico 2016-2017); recuperado el 1 de febrero de 2019, disponible en <https://fguma.cv.uma.es/course/view.php?id=2113>



# Análisis de texto con R

<http://rqda.r-forge.r-project.org/>

## What is RQDA and what are its features?

RQDA is a [R](#) package for Qualitative Data Analysis, a **free** (free as freedom) qualitative analysis software application ([BSD](#) license). It works on Windows, Linux/[FreeBSD](#) and Mac OSX platforms. RQDA is an easy to use tool to assist in the analysis of textual data. At the moment it only supports plain text formatted data. All the information is stored in a [SQLite](#) database via the R package of [RSQLite](#). The GUI is based on [RGtk2](#), via the aid of [gWidgetsRGtk2](#). It includes a number of standard Computer-Aided Qualitative Data Analysis features. In addition it seamlessly integrates with R, which means that a) statistical analysis on the coding is possible, and b) functions for data manipulation and analysis can be easily extended by writing R functions. To some extent, **RQDA and R make an integrated platform for both quantitative and qualitative data analysis.**

Through the GUI, RQDA can:

- Import documents from plain text or on-the-fly
- Import PDF highlights which is useful for content management of literature (reliant on rjpod)
- Support non-English documents. Simplified Chinese Character is well-tested under Windows as well as Debian Linux (with locale of zh\_CN.UTF-8)
- Character-level coding using codes
- File Editing after coding
- Memos of documents, codes, coding, project, files and more
- Retrieval of coding, and easily return to the original file (to ease the problem of segmentation). Conditional retrieval is supported as well.
- Single-file (\*.rqda) format, which is basically a SQLite database. Data are stored in UTF-8, so it should be portable
- Organize codes into code categories, which is key to theory building. I deliberately avoid using tree-like categorization
- Organize files into file categories
- Search files by keywords with highlighting of the keyword in the open file.
- Apply attributes to file, which is useful for content analysis
- There is a case category and related attributes of cases, which is a crucial feature in bridging qualitative and quantitative research.
- Search information about a selected case on the Internet via a pop-up menu
- Temporary delete files and codes
- Rename files, codes, code categories, and cases
- Write and organize field work journals
- And more ...


RQDA






# Análisis de texto con R

<http://iramuteq.org/>


 **SOURCEFORGE**

Open Source Software | Business Software | Services | Resources

Home / Browse / Science & Engineering / Information Analysis / iramuteq

 **iramuteq**  
Brought to you by: [pratinaud](#)

[Add a Review](#) | **Downloads: 467 This Week** | **Last Update: 2017-09-11**

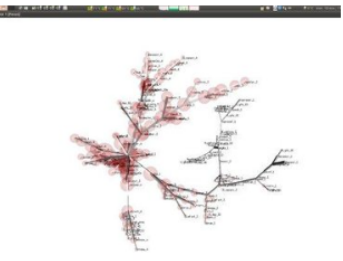
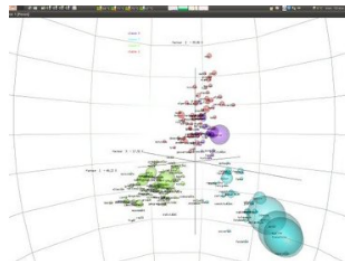


 **Download** | [Get Updates](#) | [Share This](#)

[Windows](#) | [Mac](#) | [Linux](#)

[Summary](#) | [Files](#) | [Reviews](#) | [Support](#) | [Wiki](#) | [Mailing Lists](#) | [Discussion](#)

IRAMUTEQ : Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires. Logiciel de traitement de données pour des corpus texte ou de type individus/caractères. Permet notamment de réaliser des analyses de type "ALCESTE"

### Project Samples



IRAMUTEQ

# Análisis de texto con R

<https://rtemis.hypotheses.org/>

**R.TeMiS**

**R.TeMiS**  
*Une approche intégrée et libre de l'analyse de données textuelles*

À propos   Téléchargement & Installation   Utilisation   Exportation de corpus depuis Factiva   Publications

## À propos

**R.TeMiS** [R Text Mining Solution] est un environnement graphique de travail sous R permettant de créer, manipuler et analyser des corpus de textes. Il a été conçu pour limiter les effets de « boîte noire », auxquels sont souvent confrontés les utilisateurs de logiciels de statistique lexicale, et favoriser la réflexivité dans l'usage sociologique des données textuelles.

L'architecture statistique de l'environnement **R.TeMiS** est fournie par le paquet `tm` développé par Ingo Feinerer (Feinerer, 2008; 2011; Feinerer, Hornik & Meyer, 2008). Celui-ci a été complété par d'autres paquets classiques de R comme `ca` pour la représentation des analyses factorielles des correspondances (Nenadic & Greenacre, 2007). Enfin des paquets spécifiques ont été développés pour faciliter l'usage de **R.TeMiS** dans le domaine des études sur les médias, par exemple pour la gestion des corpus constitués depuis la base de données d'articles de presse Factiva.

Afin de faciliter l'usage de **R.TeMiS** aux néo-utilisateurs de R, le développement d'un environnement graphique a été privilégié. Celui-ci se présente donc comme un menu de R Commander (Fox, 2005).

R.TeMiS est développé par [Milan Bouchet-Valat](#) (Ined) et [Gilles Bastin](#) (Sciences Po Grenoble, Pacte).

### Nouvelles de R.TeMiS

- Nouvelle version 0.7.10 22/06/2018
- Nouvelle version 0.7.9 11/07/2017
- Nouveau tutoriel pas à pas 19/05/2017
- Nouvelle version 0.7.8 02/11/2016
- Nouvelle version 0.7.7 08/07/2016

### Administration

- Connexion
- Flux [RSS](#) des articles
- [RSS](#) des commentaires
- Site de WordPress-FR

# RcmdrPlugin.temis

<https://rtemis.hypotheses.org/>

**R.TeMiS** [R Text Mining Solution] est un environnement graphique de travail sous R permettant de créer, manipuler et analyser des corpus de textes. Il a été conçu pour limiter les effets de « boîte noire », auxquels sont souvent confrontés les utilisateurs de logiciels de statistique lexicale, et favoriser la réflexivité dans l'usage sociologique des données textuelles.

Estadística lexical

tm

L'architecture statistique de l'environnement **R.TeMiS** est fournie par le paquet tm développé par Ingo Feinerer (Feinerer, 2008; 2011; Feinerer, Hornik & Meyer, 2008). Celui-ci a été complété par d'autres paquets classiques de R comme ca pour la représentation des analyses factorielles des correspondances (Nenadic & Greenacre, 2007). Enfin des paquets spécifiques ont été développés pour faciliter l'usage de **R.TeMiS** dans le domaine des études sur les médias, par exemple pour la gestion des corpus constitués depuis la base de données d'articles de presse Factiva.

ca

Factiva,  
LexisNexis  
Europresse

Afin de faciliter l'usage de **R.TeMiS** aux néo-utilisateurs de R, le développement d'un environnement graphique a été privilégié. Celui-ci se présente donc comme un menu de R Commander (Fox, 2005).

Rcmdr

Institut National  
de Études  
Démographiques

R.TeMiS est développé par Milan Bouchet-Valat (Ined) et Gilles Bastin (Sciences Po Grenoble, Pacte).

Laboratoire de  
Sciences  
 Sociales

# Statistique Textuelle

<http://ses-perso.telecom-paristech.fr/lebart/ST.html>



## Statistique Textuelle

(Ludovic Lebart et André Salem) *Préface de Christian Baudelot*

**Chapitre 0 : Préface, Sommaire, Avant Propos, Introduction** [📄 \(format pdf\)](#)

**Chapitre 1 : Domaines et problèmes** [📄 \(format pdf\)](#)

Le premier chapitre, Domaines et problèmes, évoque à la fois : les domaines disciplinaires concernés (linguistique, statistique, informatique), les pro

**Chapitre 2 : Les unités de la statistique textuelle** [📄 \(format pdf\)](#)

Le second chapitre, Les unités de la statistique textuelle, est consacré à l'étude des unités statistiques que les programmes lexicométriques devront d

**Chapitre 3 : L'analyse des correspondances** [📄 \(format pdf\)](#)

**Chapitre 4 : La classification automatique des formes et des textes** [📄 \(format pdf\)](#)

Les troisième et quatrième chapitres, L'analyse des correspondances des tableaux lexicaux, et La classification automatique des formes et des textes,

**Chapitre 5 : Typologies, visualisations** [📄 \(format pdf\)](#)

Le cinquième chapitre : Typologies, visualisations, applique les outils présentés aux chapitres trois et quatre à la description des associations entre fi  
méthodologique.

**Chapitre 6 : Eléments caractéristiques, réponses ou textes modaux** [📄 \(format pdf\)](#)

Pour compléter ces représentations synthétiques, le sixième chapitre, Eléments caractéristiques, réponses ou textes modaux, présente les calculs dits  
permet de replacer les formes dans leur contexte, et de caractériser, lorsque cela est possible, des parties de texte, en général volumineuses, par des p  
le corpus, choisies en raison de leur caractère représentatif.

**Chapitre 7 : Partitions longitudinales, contiguité** [📄 \(format pdf\)](#)

Le septième chapitre, Partitions longitudinales, contiguité, traite le problème des informations a priori qui concernent les parties d'un corpus. Dans d  
(séries textuelles chronologiques). On étudie dans ce chapitre, en présentant une méthode et de nombreux exemples d'application, les relations de dé

**Chapitre 8 : Analyse discriminante textuelle** [📄 \(format pdf\)](#)

Le huitième chapitre, consacré à l'Analyse discriminante textuelle, étudie, au sens statistique du terme, le pouvoir de discrimination des textes. Com  
base de données textuelles ? On tente dans ce chapitre, qui contient des exemples d'application variés, de montrer quels sont les apports de la statisti

**Annexes et fin :** [📄 \(format pdf\)](#)

Annexe A Description sommaire de quatre logiciels  
Annexe B Esquisse des algorithmes et structures de données  
Glossaire  
Références bibliographiques  
Index des auteurs

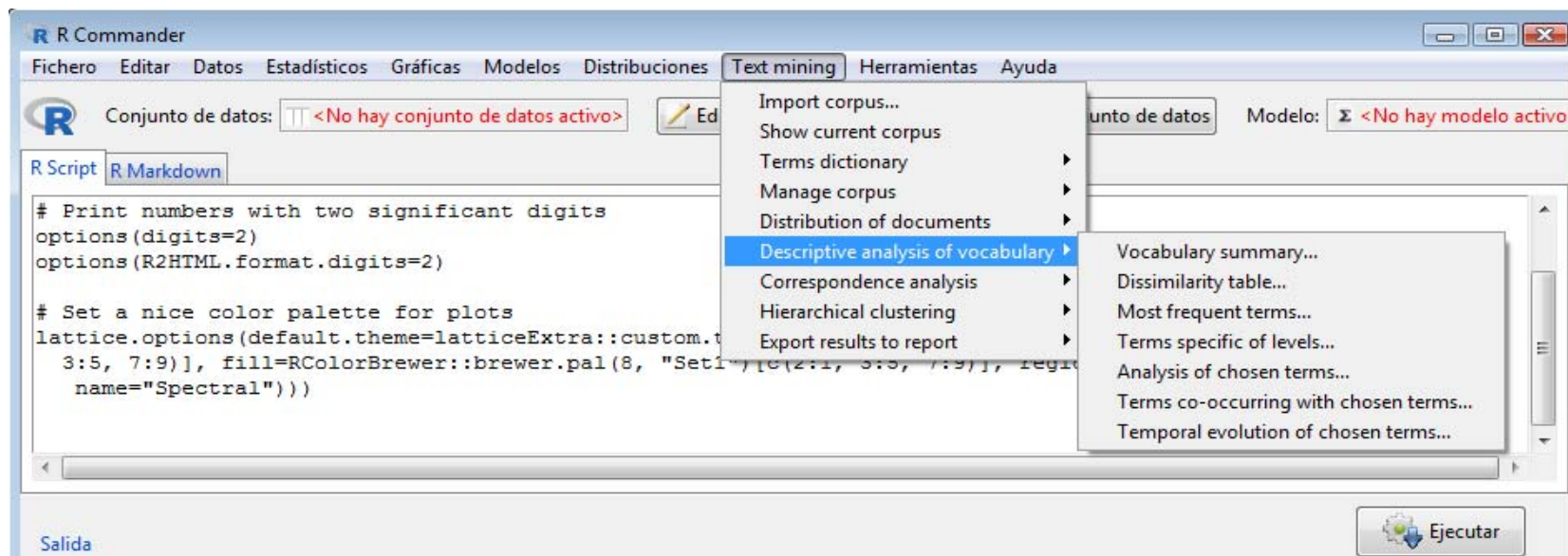
## *R Text Mining Solution (R.TeMiS)*

Bouchet-Valat, M., & Bastin, G. (2013). RcmdrPlugin.temis, a Graphical Integrated Text Mining Solution in R. *The R Journal* , 188-196.

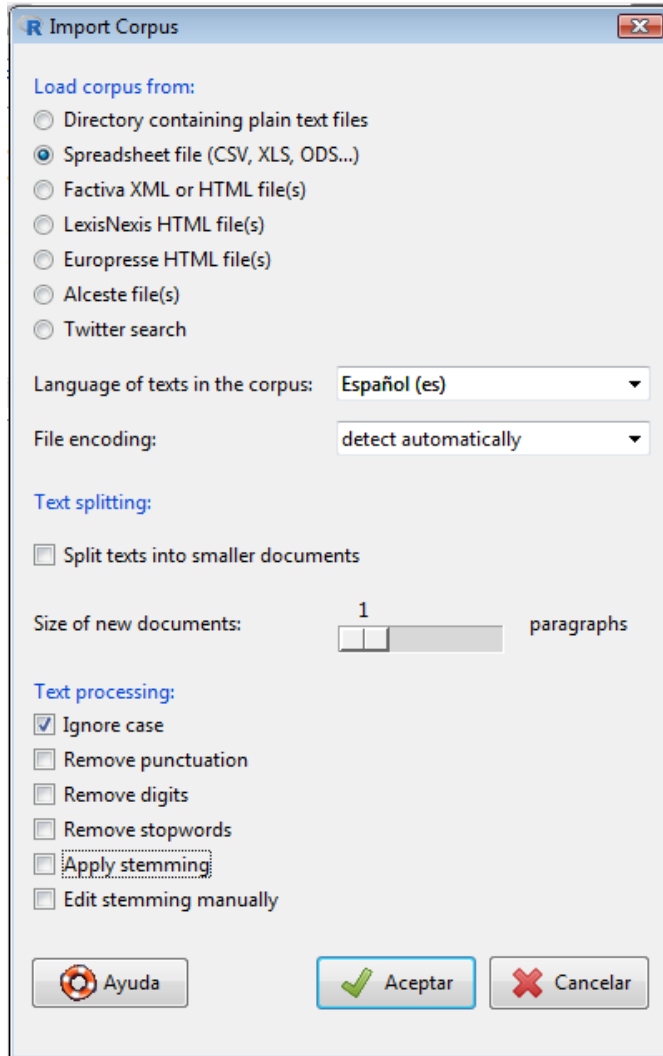
Bouchet-Valat, M. (2016). *Package RcmdrPlugin.temis*.

<https://cran.r-project.org/web/packages/RcmdrPlugin.temis/>

Garnier, B. (2014). *R.TeMiS. Une approche intégrée et libre de l'analyse de données textuelles*. <http://rtemis.hypotheses.org/>



## *R Text Mining Solution (R.TeMiS)*



The screenshot shows the 'Import Corpus' dialog box in the R.TeMiS application. It has a title bar with the R logo and the text 'Import Corpus'. The dialog is organized into several sections:

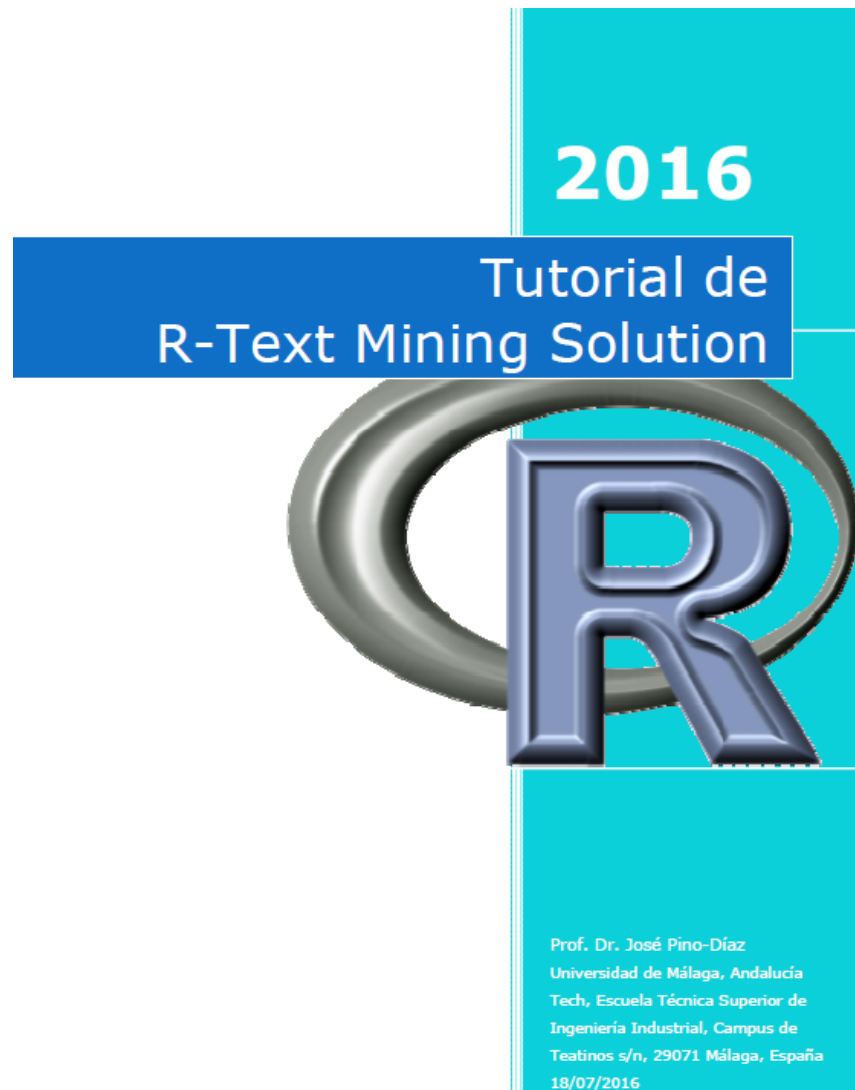
- Load corpus from:** A group of radio buttons with the following options: 'Directory containing plain text files', 'Spreadsheet file (CSV, XLS, ODS,...)' (which is selected), 'Factiva XML or HTML file(s)', 'LexisNexis HTML file(s)', 'Europresse HTML file(s)', 'Alceste file(s)', and 'Twitter search'.
- Language of texts in the corpus:** A dropdown menu currently showing 'Español (es)'.
- File encoding:** A dropdown menu currently showing 'detect automatically'.
- Text splitting:** A checkbox for 'Split texts into smaller documents' is unchecked. Below it, 'Size of new documents:' is set to '1' with a slider bar, followed by the unit 'paragraphs'.
- Text processing:** A group of checkboxes: 'Ignore case' (checked), 'Remove punctuation' (unchecked), 'Remove digits' (unchecked), 'Remove stopwords' (unchecked), 'Apply stemming' (unchecked), and 'Edit stemming manually' (unchecked).

At the bottom of the dialog are three buttons: 'Ayuda' (Help) with a question mark icon, 'Aceptar' (Accept) with a green checkmark icon, and 'Cancelar' (Cancel) with a red X icon.

1. Visualización del corpus activo y de los diccionarios de términos.
2. Gestión y distribución del corpus.
  1. Creación y análisis de subcorpus por términos o por variables
  2. Creación de tablas y gráficos de distribución.
3. Análisis descriptivo del léxico.
  1. Resumen cuantitativo del vocabulario de términos.
  2. Tabla de disimilaridad.
  3. Términos más frecuentes.
  4. Términos específicos por modalidades de la variable.
  5. Análisis de términos concretos.
  6. Términos que coocurren con otros concretos.
  7. Evolución temporal de términos concretos.
4. Análisis factorial de correspondencias.
  1. AFC de la matriz documentos-términos sin agregar ninguna variable.
  2. AFC de la tabla lexical completa agregando variables.
5. Clasificación ascendente jerárquica.



# Tutorial R.TeMiS (Pino, 2016)



## Contenido

Introducción .....	4
Comenzando a trabajar con R.TeMiS .....	5
Importar Corpus .....	8
Visualización del corpus activo y de los diccionarios de términos .....	13
Visualizar el corpus de documentos.....	13
Visualizar el diccionario de términos.....	14
Gestión y distribución del corpus.....	15
Análisis descriptivo del léxico.....	21
Resumen cuantitativo del vocabulario de términos. ....	22
Tabla de disimilaridad.....	23
Términos más frecuentes.....	24
Términos específicos por modalidades de la variable. ....	25
Análisis de términos concretos. ....	27
Términos que coocurren con otros concretos.....	29
Evolución temporal de términos concretos. ....	29
Análisis de correspondencias aplicado a un corpus de documentos .....	30
Concepto de similitud, disimilaridad, distancia y proximidad entre documentos. ....	30
Análisis de correspondencias.....	30
Análisis de correspondencias con RTemis. ....	31
Interpretación del diagrama de correspondencias .....	34
Procedimiento de análisis de correspondencias con RTemis. ....	35
I) AFC de la matriz documentos-términos sin agregar ninguna variable ...	35
II) AFC de la tabla lexical completa agregando variables ( <i>full document-term matrix by variables</i> ).....	40
Clasificación ascendente jerárquica aplicada a un corpus de documentos .....	45
Concepto de similitud y distancia entre documentos. ....	45
Concepto de clasificación jerárquica. ....	45
Clasificación ascendente jerárquica.....	46
Clasificación ascendente jerárquica con RTemis. ....	47

## Algunos ejemplos de aplicación de R.Temis

- Análisis del texto de Menéndez Pelayo “La Ciencia. 1ª parte, Al Sr. D. Gumersindo Laverde”
- KDD con R.TeMiS, su aplicación en el Proyecto Exhibitium
- Estudio de la relación de asociación entre operadores de telefonía, tipos de red, potencia de la señal recibida y modelo de móvil, según el conjunto de datos públicos “Málaga CitySense - Mayo 2015”.
- “*Reliability and Failure\**”, patentes del año 2018 en la BD Espacenet.
- Análisis de respuestas de encuestas abiertas en el MOOC “Innovación en el sector público”.

# Análisis del texto de Menéndez Pelayo

## “La Ciencia. 1ª parte, Al Sr. D. Gumersindo Laverde”

Estadística\_textual

Europress

Factiva

Fiabilidad\_ICYT\_2014

LexisNexis

**MENÉNDEZ\_PELAYO\_LA\_CIENCIA\_**

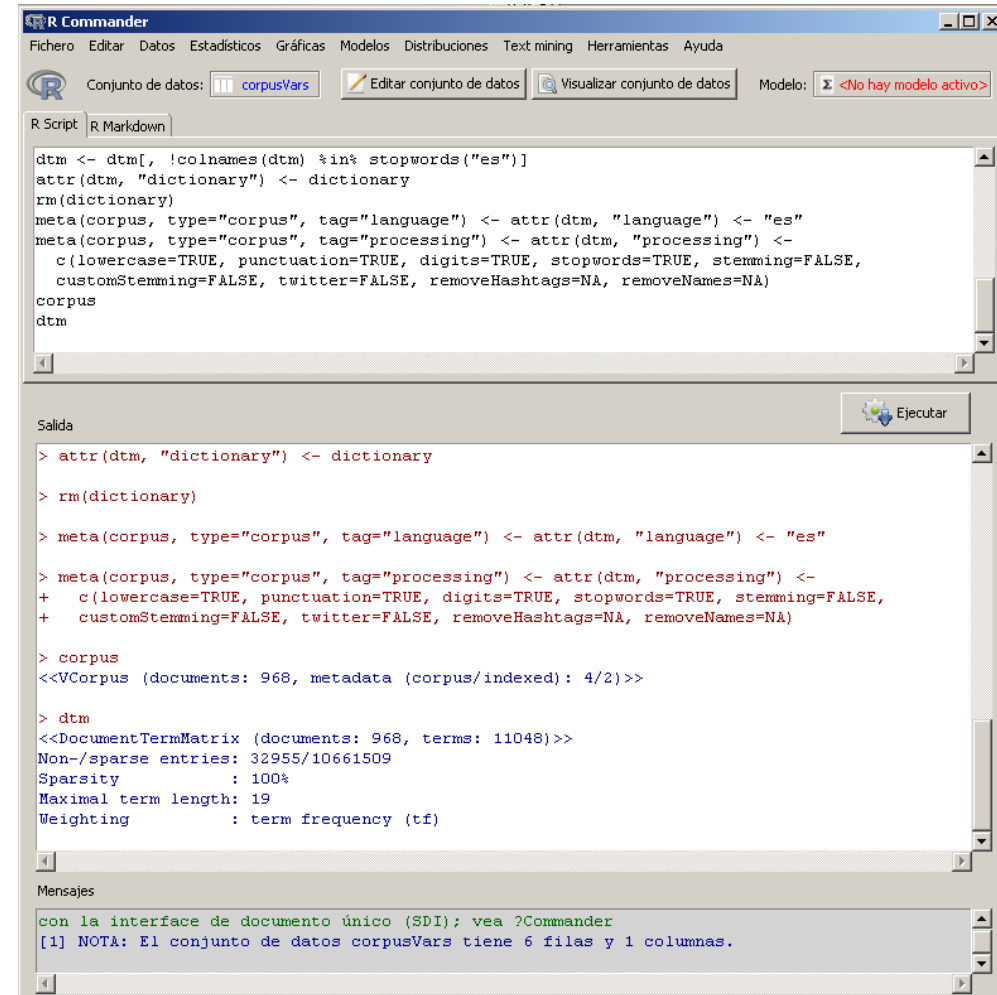
Prueba\_análisis\_R\_BD\_CSIC

PUBMED

RTemis\_AAI

### Nombre

- 1.MENÉNDEZ\_PELAYO\_ACTIVIDAD\_INTELCTUAL\_ESPAÑA
- 2.MENÉNDEZ PELAYO. DE RE BIBLIOGRAPHICA
- 4.MENÉNDEZ PELAYO. MR. MASSON REDIVIVO
- 5.MENÉNDEZ PELAYO. MONOGRAFÍAS EXPOSITIVO-CRÍTICAS
- 6.MENÉNDEZ PELAYO. PROSÍGUESE EL PENSAMIENTO DE LAS CARTAS ANTERIORES
- 8.MENÉNDEZ PELAYO. MR. MASSON REDIMUERTO (SEGUNDA CONTESTACIÓN A D. M...



The screenshot shows the R Commander window with the following content:

**R Commander**  
Fichero Editar Datos Estadísticos Gráficas Modelos Distribuciones Text mining Herramientas Ayuda

Conjunto de datos: corpusVars [Editar conjunto de datos] [Visualizar conjunto de datos] Modelo: <No hay modelo activo>

**R Script** | R Markdown

```
dtm <- dtm[, !colnames(dtm) %in% stopwords("es")]
attr(dtm, "dictionary") <- dictionary
rm(dictionary)
meta(corpus, type="corpus", tag="language") <- attr(dtm, "language") <- "es"
meta(corpus, type="corpus", tag="processing") <- attr(dtm, "processing") <-
  c(lowercase=TRUE, punctuation=TRUE, digits=TRUE, stopwords=TRUE, stemming=FALSE,
    customStemming=FALSE, twitter=FALSE, removeHashtags=NA, removeNames=NA)
corpus
dtm
```

**Salida** [Ejecutar]

```
> attr(dtm, "dictionary") <- dictionary
> rm(dictionary)
> meta(corpus, type="corpus", tag="language") <- attr(dtm, "language") <- "es"
> meta(corpus, type="corpus", tag="processing") <- attr(dtm, "processing") <-
+ c(lowercase=TRUE, punctuation=TRUE, digits=TRUE, stopwords=TRUE, stemming=FALSE,
+ customStemming=FALSE, twitter=FALSE, removeHashtags=NA, removeNames=NA)
> corpus
<<VCorpus (documents: 968, metadata (corpus/indexed): 4/2)>>
> dtm
<<DocumentTermMatrix (documents: 968, terms: 11048)>>
Non-/sparse entries: 32955/10661509
Sparsity           : 100%
Maximal term length: 19
Weighting           : term frequency (tf)
```

**Mensajes**

```
con la interface de documento único (SDI); vea ?Commander
[1] NOTA: El conjunto de datos corpusVars tiene 6 filas y 1 columnas.
```

### Upper part of documents dendrogram



Hierarchical clustering

Hierarchical clustering of 906 documents using 397 terms (Ward's method with Chi-squared distance).

Clusters summary:

	1	2	3	4	5	6	7	8	9
Number of documents	195.0	441.2	6.00	5.00	78.0	131.1	3.00	9.00	2.00
% of documents	21.5	48.7	0.66	0.55	8.6	14.5	0.33	0.99	0.22
Within-cluster variance	15.1	46.7	0.00	0.00	63.0	67.0	0.00	45.12	0.00
	10	11	12	13	14	15			
Number of documents	2.00	2.00	2.00	7.00	6.00	17.0			
% of documents	0.22	0.22	0.22	0.77	0.66	1.9			
Within-cluster variance	0.00	0.00	0.00	53.40	53.57	65.7			

Terms specific of cluster 1:

	% Term/Level	% Level/Term	Global %	Level Global	t	value	Prob.
revilla	0.5701	92	0.420	137	149	7.0	0
tan	0.4328	92	0.318	104	113	6.1	0
si	0.5493	88	0.423	132	150	5.7	0
sino	0.2788	93	0.203	67	72	5.0	0
pues	0.2039	94	0.146	49	52	4.5	0
edición	0.1124	20	0.375	27	133	-11.2	0
nota	0.0791	22	0.245	19	87	-8.7	0
vida	0.0749	22	0.231	18	82	-8.4	0
tercera	0.0832	26	0.217	20	77	-7.4	0
tomo	0.0541	22	0.169	13	60	-7.2	0
séneca	0.0291	18	0.113	7	40	-6.4	0
madrid	0.0666	27	0.166	16	59	-6.3	0
et	0.1290	37	0.234	31	83	-5.6	0
biografía	0.0166	15	0.076	4	27	-3.5	0
discurso	0.0499	29	0.118	12	42	-5.0	0
berlín	0.0083	1	0.054	2	19	-4.9	0
José	0.0624	33	0.130	15	46	-4.7	0
quijote	0.0000	0	0.031	0	11	-4.5	0
traducción	0.0000	0	0.031	0	11	-4.5	0
véase	0.0000	0	0.031	0	11	-4.5	0

Documents specific of cluster 1:

	Chi2 dist.	to centroid
1.MENÉNDEZ_PELAYO_ACTIVIDAD_INTELLECTUAL_ESPAÑA.txt.22		24.2
2.MENÉNDEZ_PELAYO_DE RE BIBLIOGRAPHICA.txt.68		28.2
1.MENÉNDEZ_PELAYO_ACTIVIDAD_INTELLECTUAL_ESPAÑA.txt.18		28.9
4.MENÉNDEZ_PELAYO_MR. MARSON REDIVIVO.txt.19		30.6
2.MENÉNDEZ_PELAYO_DE RE BIBLIOGRAPHICA.txt.16		31.6

1.MENÉNDEZ\_PELAYO\_ACTIVIDAD\_INTELLECTUAL\_ESPAÑA.txt.22

Pero se dirá ¿por qué obtuvieron tan escaso florecimiento las [p. 48] ciencias exactas, físicas y naturales , sino por la rigidez con que el Estado negó siempre la libertad de la ciencia?

Entendámonos: en primer lugar, niego el supuesto formulado en términos tan absolutos: verdad es que no apareció en España ningún Galileo, Descartes, Newton, Lagrange, Lavoisier o Linneo; confieso de buen grado nuestra inferioridad en esta parte, no lo da Dios todo a todos; quizá el terreno no estaba tan bien preparado; quizá la genialidad española no tra tanto por ese camino como por otros, quizá la época en que España fue grande y sabia no coincidió con la madurez sino con los primeros ensayos y tentativas del genio analítico y experimental, pero es lo cierto que en esos oscuros siglos debieron las ciencias de la naturaleza considerables adelantos a muchos españoles, acaudaláronse la Zoología y la Botánica con innumerables noticias sobre la fauna y la flora de los países americanos, espardidas en los libros de Gonzalo Fernández de Oviedo y otros primitivos historiadores de Indias, y luego más científicamente expuestas en los tratados de Nicolás Monardes, Francisco Hernández y José de Acosta, brillaron Quer, Gómez Ortega, Cavanilles y tantos otros sabios ilustradores del reino vegetal , de que en su laureada obra La Botánica y los Botánicos de la Península da cumplida noticia el Sr. D. Miguel Colmeiro; hicieron importantes estudios sobre los metales Álvaro Alonso Barba, Bernal Pérez de Vargas y otros menos conocidos autores; publicáronse notables comentarios y traducciones de Aristóteles y Teofrasto, de Arquímedes y Euclides, de Dioscórides y Plinio; no faltaron matemáticos y físicos tan memorables como Núñez, inventor del nonus, el docto humanista Fernán Pérez de Oliva, que escribió De Magnetis y se amasó en bulto modo de que cada ciencia indaga comunicase dos cuantos [11] el compendio Teófil

Clusters summary

Cluster 1:

1.MENÉNDEZ\_PELAYO\_A

2.MENÉNDEZ\_PELAYO\_DI

1.MENÉNDEZ\_PELAYO\_A

4.MENÉNDEZ\_PELAYO\_MI

2.MENÉNDEZ\_PELAYO\_DI

Cluster 2:

1.MENÉNDEZ\_PELAYO\_A

5.MENÉNDEZ\_PELAYO\_FF

5.MENÉNDEZ\_PELAYO\_FF

5.MENÉNDEZ\_PELAYO\_FF

5.MENÉNDEZ\_PELAYO\_FF

6.MENÉNDEZ\_PELAYO\_FF

Cluster 3:

5.MENÉNDEZ\_PELAYO\_MI

6.MENÉNDEZ\_PELAYO\_FF

2.MENÉNDEZ\_PELAYO\_DI

4.MENÉNDEZ\_PELAYO\_MI

6.MENÉNDEZ\_PELAYO\_MI

Cluster 4:

5.MENÉNDEZ\_PELAYO\_MI

5.MENÉNDEZ\_PELAYO\_MI

5.MENÉNDEZ\_PELAYO\_MI

5.MENÉNDEZ\_PELAYO\_MI

5.MENÉNDEZ\_PELAYO\_MI

Cluster 5:

5.MENÉNDEZ\_PELAYO\_MI

4.MENÉNDEZ\_PELAYO\_FF

4.MENÉNDEZ\_PELAYO\_MI

5.MENÉNDEZ\_PELAYO\_MI

5.MENÉNDEZ\_PELAYO\_MI

Cluster 6:

5.MENÉNDEZ\_PELAYO\_MI

5.MENÉNDEZ\_PELAYO\_MI

5.MENÉNDEZ\_PELAYO\_MI

5.MENÉNDEZ\_PELAYO\_DI

5.MENÉNDEZ\_PELAYO\_MI

Cluster 7:

2.MENÉNDEZ\_PELAYO\_DI

2.MENÉNDEZ\_PELAYO\_DI

5.MENÉNDEZ\_PELAYO\_MI

Cluster 8:

1.MENÉNDEZ\_PELAYO\_A

1.MENÉNDEZ\_PELAYO\_A

1.MENÉNDEZ\_PELAYO\_A

Cluster 9:

5.MENÉNDEZ\_PELAYO\_DI

5.MENÉNDEZ\_PELAYO\_MI

Cluster 10:

5.MENÉNDEZ\_PELAYO\_MI

5.MENÉNDEZ\_PELAYO\_MI

Cluster 11:

5.MENÉNDEZ\_PELAYO\_MI

5.MENÉNDEZ\_PELAYO\_MI

Cluster 12:

2.MENÉNDEZ\_PELAYO\_DI

5.MENÉNDEZ\_PELAYO\_MI

Cluster 13:

8.MENÉNDEZ\_PELAYO\_MI

5.MENÉNDEZ\_PELAYO\_MI

5.MENÉNDEZ\_PELAYO\_MI

5.MENÉNDEZ\_PELAYO\_MI

Cluster 14:

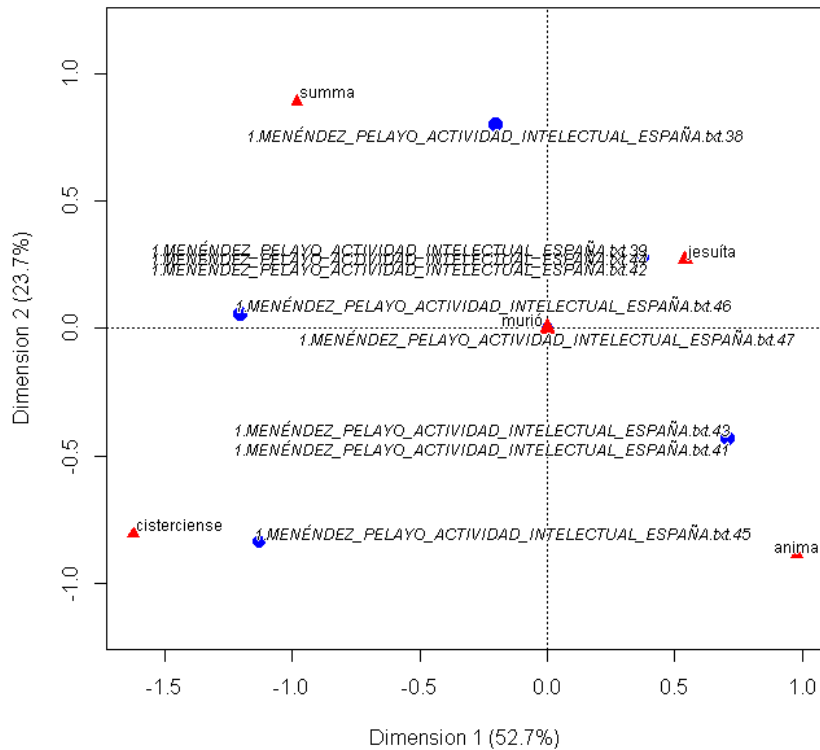
5.MENÉNDEZ\_PELAYO\_DI

5.MENÉNDEZ\_PELAYO\_DI

1.MENÉNDEZ\_PELAYO\_A

1.MENÉNDEZ\_PELAYO\_A

Análisis del texto de Menéndez Pelayo  
“La Ciencia. 1ª parte, Al Sr. D. Gumersindo Laverde”



Correspondence analysis of 1 active variable(s) (aggregating 906 documents), 597 terms and 2 supplementary variable(s).

**Axes summary:**

	Axis				
	1	2	3	4	5
Inertia (%)	18	14	8.8	8.4	7.7
Cumulated inertia (%)	18	32	41.3	49.7	57.3

**Most contributive terms on negative side of axis 1:**

	Position	Contribution (%)	Quality (%)
murió	-0.094	0.00136	0.00262
sr	-0.019	0.00125	0.92009
historia	-0.019	0.00094	0.40885
jesuita	-0.071	0.00079	0.00342
españa	-0.019	0.00078	3.10940
revilla	-0.019	0.00068	0.12332
si	-0.019	0.00067	0.16597
et	-0.025	0.00067	0.00534
sígl	-0.019	0.00063	0.44544
filosofía	-0.019	0.00061	0.56409
obras	-0.020	0.00061	0.41057
tan	-0.019	0.00055	0.30011
ciencia	-0.019	0.00054	0.23751
in	-0.032	0.00042	0.00197
tomo	-0.018	0.00026	0.02540
berlín	-0.032	0.00024	0.00984
volumenes	-0.025	0.00020	0.02277
der	-0.033	0.00019	0.00073
escribió	-0.029	0.00015	0.02571
cuestiones	-0.025	0.00013	0.02611
philosophia	-0.023	0.00012	0.02884
comentarios	-0.031	0.00012	0.01584
cuatro	-0.025	0.00012	0.02404
insigne	-0.028	0.00011	0.01546

Active levels on negative side of axis 1:

	Position	Contribution (%)	Quality (%)
Cluster 1	-0.014	0.024491	0.214637
Cluster 6	-0.033	0.007786	0.029864
Cluster 2	-0.011	0.006198	0.032390
Cluster 5	-0.034	0.003108	0.011928
Cluster 8	-0.084	0.002390	0.003026
Cluster 13	-0.044	0.000370	0.000768
Cluster 14	-0.003	0.000163	0.000063
Cluster 15	-0.030	0.000244	0.001221
Cluster 4	-0.035	0.000080	0.000069
Cluster 7	-0.039	0.000058	0.000126
Cluster 10	-0.045	0.000053	0.000159
Cluster 9	-0.039	0.000039	0.000108
Cluster 11	-0.037	0.000035	0.000124
Cluster 12	-0.031	0.000025	0.000093

**Most extreme documents on negative side of axis 1:**

	Position	Quality (%)
1.MENÉNDEZ_PELAYO_ACTIVIDAD_INTELLECTUAL_ESPAÑA.txt.38	-0.117	0.00211
1.MENÉNDEZ_PELAYO_ACTIVIDAD_INTELLECTUAL_ESPAÑA.txt.42	-0.117	0.00211

Access summary

Axis 1 - Negative Side

1.MENÉZES\_PELAYO\_A  
1.MENÉZES\_PELAYO\_A  
1.MENÉZES\_PELAYO\_A  
1.MENÉZES\_PELAYO\_A  
1.MENÉZES\_PELAYO\_A  
1.MENÉZES\_PELAYO\_A  
1.MENÉZES\_PELAYO\_A  
2.MENÉZES\_PELAYO\_DI  
1.MENÉZES\_PELAYO\_DI  
1.MENÉZES\_PELAYO\_DI  
5.MENÉZES\_PELAYO\_M  
5.MENÉZES\_PELAYO\_M  
5.MENÉZES\_PELAYO\_M  
5.MENÉZES\_PELAYO\_M

Axis 1 - Positive Side

2.MENÉZES\_PELAYO\_DI  
2.MENÉZES\_PELAYO\_DI  
4.MENÉZES\_PELAYO\_M  
5.MENÉZES\_PELAYO\_M  
6.MENÉZES\_PELAYO\_P  
8.MENÉZES\_PELAYO\_M  
2.MENÉZES\_PELAYO\_DI  
2.MENÉZES\_PELAYO\_DI  
2.MENÉZES\_PELAYO\_DI

Axis 1 - Variables

Axis 2 - Negative Side

1.MENÉZES\_PELAYO\_A  
1.MENÉZES\_PELAYO\_A  
1.MENÉZES\_PELAYO\_A  
1.MENÉZES\_PELAYO\_A  
1.MENÉZES\_PELAYO\_A  
1.MENÉZES\_PELAYO\_A  
1.MENÉZES\_PELAYO\_A  
5.MENÉZES\_PELAYO\_M  
1.MENÉZES\_PELAYO\_A  
1.MENÉZES\_PELAYO\_A  
1.MENÉZES\_PELAYO\_A  
8.MENÉZES\_PELAYO\_M  
2.MENÉZES\_PELAYO\_DI  
2.MENÉZES\_PELAYO\_DI  
2.MENÉZES\_PELAYO\_DI  
2.MENÉZES\_PELAYO\_DI  
5.MENÉZES\_PELAYO\_M  
2.MENÉZES\_PELAYO\_DI  
2.MENÉZES\_PELAYO\_DI  
6.MENÉZES\_PELAYO\_M  
5.MENÉZES\_PELAYO\_M  
5.MENÉZES\_PELAYO\_M  
2.MENÉZES\_PELAYO\_DI  
2.MENÉZES\_PELAYO\_DI  
2.MENÉZES\_PELAYO\_DI  
5.MENÉZES\_PELAYO\_M  
5.MENÉZES\_PELAYO\_M

Axis 2 - Positive Side

2.MENÉZES\_PELAYO\_DI  
2.MENÉZES\_PELAYO\_DI  
2.MENÉZES\_PELAYO\_DI  
2.MENÉZES\_PELAYO\_DI  
5.MENÉZES\_PELAYO\_M  
2.MENÉZES\_PELAYO\_DI  
2.MENÉZES\_PELAYO\_DI  
6.MENÉZES\_PELAYO\_M  
5.MENÉZES\_PELAYO\_M  
5.MENÉZES\_PELAYO\_M  
2.MENÉZES\_PELAYO\_DI  
2.MENÉZES\_PELAYO\_DI  
2.MENÉZES\_PELAYO\_DI  
5.MENÉZES\_PELAYO\_M  
5.MENÉZES\_PELAYO\_M

Axis 2 - Variables

# KDD con R.TeMiS, su aplicación en el Proyecto Exhibitium



The image shows a screenshot of the Exhibitium website. The background is a dark, textured red with a large, realistic red apple on the right side. At the top, there is a navigation bar with links: Inicio, Proyecto, Objetivos, Documentación, Equipo, Agenda, Blog, Contacto, and a search icon. The logo 'EXHIBITIUM' is in the top right corner. The main heading in white text reads: 'EXHIBITIUM ES UN PROYECTO DE INVESTIGACIÓN FINANCIADO POR LA FUNDACIÓN BBVA'. Below this, a yellow box contains the following information:

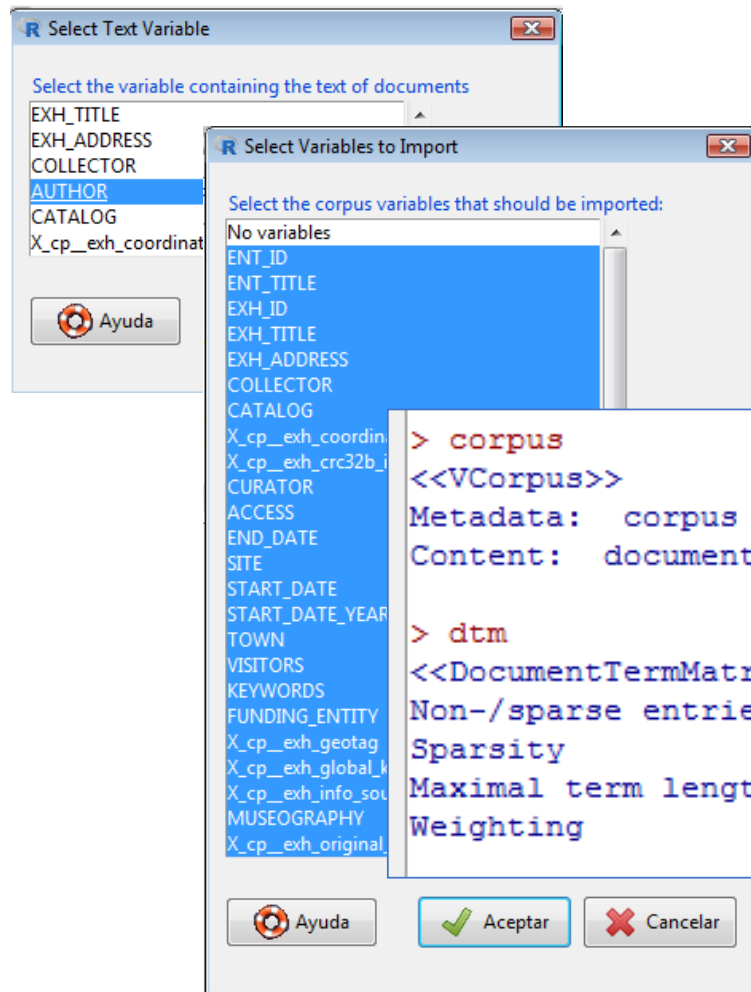
**Proyecto:** Generación de conocimiento sobre exposiciones artísticas temporales para su reutilización y aprovechamiento multivalente.  
**Investigadora principal:** Nuria Rodríguez Ortega (Universidad de Málaga)  
**Grupo de investigación:** iArtHis\_Lab  
**Categoría:** Humanidades Digitales  
**Ayuda Fundación BBVA:** 60.000 Euros  
**Web:** <http://exhibitium.com/>



# Conjunto de datos “Expofinder” 3391 exposiciones

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
1	ENT_ID	ENT_TITL	EXH_ID	EXH_TITL	cp_exh	cp_exh	cp_exh	cp_exh	cp_exh	cp_exh	cp_exh	cp_exh	cp_exh	cp_exh	cp_exh	cp_exh	cp_exh	cp_exh
2	1040	Museo de	68757	Bartolozzi	Calle de Santo Domín	67322: Ra	68748: Ba	42.8194188	-1.646151	68728: Ra	Gratis par	###	Museo de	###	Pamplona	< 100		
3	1075	Museo Ev	25042	Vagabundo	Camino Cabueñes, 2	25404: Evaristo Valle	43.533583	3fd01aa1			Sistema n	###	Museo Ev	###	Gijón: Astu	< 100	arte	
4	1075	Museo Ev	58543	Forma y O	Camino Cabueñes, 2	23241: Consuelo Valli	43.533583	0a0fab7e			No dispon	###	Museo Ev	###	Gijón: Astu	< 100	mus	
5	1075	Museo Ev	65636	LIQUID, F	Camino Cabueñes, 3	67624: Michael Dunev	43.534038	434c08a9			No dispon	###	Museo Ev	###	Gijón: Astu	< 100	mus	
6	1075	Museo Ev	65638	Faena de	Camino C	1209: Mus	25404: Evaristo Valle	43.533944	4b522248	67626: Fra	No dispon	###	Museo Ev	###	Gijón: Astu	< 100	mus	
7	1077	Museo Ba	47777	El oficio d	Calle Trinidad, 17, 33	25459: César Lucas	43.5441507	-5.663827	29999998,0		No dispon	###	Museo Ba	###	Gijón: Astu	< 100		
8	1077	Museo Ba	49504	Silencios	Calle Trinidad, 17, 33	49505: Mingo Venero	43.5441507	-5.663827	49510: Sa		No dispon	###	Museo Ba	###	Gijón: Astu	< 100		
9	1085	Museo Etr	32887	Exposició	LLN-8, 132, 33509 Lla	33074: Cordelia Pickfo	43.412467	9fc6a0dc			No dispon	###	Centro Cu	###	Porrua: Ag	< 100	arte	
10	1156	Museo Mu	21523	27 negros	Rúa de Ramón del Va	24337: Antonio Gonzé	42.883074	8d527b57			Gratis par	###	Centro Ga	###	Santiago	< 100	arte	
11	1156	Museo Mu	22068	Transform	Malloy Abaixo, 2, 157	24566: Vic	24592: Vic	42.892179	db0052d9	24586: Ali	No dispon	###	Centro Ga	###	Santiago	< 100	arte	
12	1156	Museo Mu	22934	De las tier	Praza Maior, 2, 32005	25731: Natasha Leler	42.335789	00ac69bc			No dispon	2014-08-1	Museo Etr	###	Ourense	< 100	arte	
13	1156	Museo Mu	23062	Exposició	Rúa do Vilar, 49-51, 1	22113: Cristina Iglesias	42.878213	5bf2ffd6	25349: Mi		No dispon	###	Centro Ga	###	Santiago	< 100	arte	
14	1156	Museo Mu	24071	Antonio M	Rúa Montero Ríos, 18	27022: Antonio Massó	42.240531	5d16139b	27025: Co		Gratis par	###	Museo Ma	###	Ponteved	< 100	arte	
15	1156	Museo Mu	24112	La conser	LU-P-1102, 6, 27259 O	23812: Colectivo sin d	43.161734	0c476378			No dispon	###	Museo Ar	###	Castro de	< 100	exp	
16	1156	Museo Mu	24113	Il Cammin	Plaza Puerta del Sol, 1	27028: Steve McCurry		5192cab1			Gratis par	###	Museo da	###	Santiago	< 100	arte	
17	1156	Museo Mu	24983	James W	Rúa de Ramón del Va	26772: James Wellino	42.883074	ec208b86	27140: Ant		No dispon	###	Centro Ga	###	Santiago	< 100	arte	
18	1156	Museo Mu	24984	Curculio B	Rúa de Ramón del Va	26393: Mark Manders	42.883074	e2f16a38	26394: Jav		Gratis par	###	Centro Ga	###	Santiago	< 100	arte	
19	1156	Museo Mu	25123	A orde da	Rúa de Ramón del Va	27185: Carlos León	42.883074	bc88dfc3	27189: An		Gratis par	###	Centro Ga	###	Santiago	< 100	arti	
20	1156	Museo Mu	25551	Océano	Av. da Atlántida, 160, 1	25690: Antón Patiño	42.224067	f514e7f8			No dispon	###	Museo do	###	Vigo: Gali	< 100	arti	
21	1156	Museo Mu	33188	La Ría de	Praza de	28487: Ger	28487: Germán Cruce	42.738117	790b67e9		No dispon	###	Casa de la	###	Padrón: G	< 100	arti	
22	1156	Museo Mu	43294	La Reboti	Rúa Padre Amodeo	22108: Eduardo Chillí	42.433485	91da5cfe	43691: Co		No dispon	###	Museo de	###	Ponteved	< 100	pint	
23	116	Museu Ar	50171	Cantigas	Carrer Sant Bartolomé	51168: Juan Fernánde	38.484640	f3d370ad			No dispon	###	Museu Ar	###	Petrer: Va	< 100	arti	
24	116	Museu Ar	61486	Made in A	Av. del Almte, Julio G	55293: [Sin determina	38.340201	82d4346b	63824: Da		Gratis par	###	Lonja del	###	Alicante: V	< 100	mus	
25	1178	Museo Art	21695	Coco Cha	Calle Patio Chico, 2, 1	25653: Alberto Giac	40.959626	9e2745dd	29256: Ma		No dispon	###	Museo Art	###	Salamanca	100 a 499	pint	
26	1197	Museo Ar	7622	Cerámica	Calle Dr. Cervia Cabre	23812: Colectivo sin d	28.463812	-16.249057	99999999,0		No dispon	###	Museo de	###	Tenerife	< 100		
27	1198	Museo de	7626	Figuras p	Calle Dr. Cervia Cabre	7629: Christine Mathie	28.463812	-16.249057	99999999,0		Gratis par	###	Museo de	###	Tenerife	< 100		
28	1199	Centro de	25102	Collages	Avenida de San Seb	25120: Jorge Rodrígu	28.463852	90da2094	25121: Isi		No dispon	###	Tea Tene	###	Santa Cru	< 100	arte	
29	1201	TEA Tene	21252	Waraos	Av. de San Sebastián	81994: Beatriz Rodrí	28.463851	b0bb2070			Gratis par	###	TEA Tene	###	Tenerife	< 100	arte	
30	1209	Museo de	48036	Sotileza /	Calle Rampa Sotileza	29276: Concha García	43.459909	f3d370ad	49387: Sa		Gratis par	###	Obra al ai	###	Santande	< 100	arte	
31	1210	Museo de	56733	Una arque	Calle Bailén, 4, 39004	55293: [Sin determina	43.462253	2cc921a3	69578: Ma		No dispon	###	Museo de	###	Santande	< 100	mus	
32	1212	Museo Na	56209	De Bison	CA-134, 39360 Santilla	23812: Colectivo sin d	43.377402	a48d0144			No dispon	###	Museo Na	###	Santillana	< 100	arte	
33	1212	Museo Na	62965	Elle n'eta	CA-134, 39360 Santilla	63056: Sara Morante	43.377427	76238111	-4,12239505		No dispon	###	Museo Na	###	Santillana	< 100		
34	1223	Museo de	33197	La Real F	Paseo Pocillo, 1, 40100 La Granja de S. Ildel	40.902871	2dfe7a50				No dispon	###	Museo Te	###	Real Sitio	< 100	arte	
35	1240	Museo de	64312	El compá	Calle Manuel Rojas M	64910: Ángel Idígoras	37.3889387	-5.991232	2999999979,1		No dispon	###	Museo de	###	Sevilla: Ar	< 100		
36	1240	Museo de	64313	Camino de	Calle Manuel Rojas M	64384: Mauro Di Girol	37.3889387	-5.991232	2999999979,1		No dispon	###	Museo de	###	Sevilla: Ar	< 100		
37	136	Museo Ca	45185	Selección	Calle Navarro Rodrí	81256: Rafael Gadea	36.839738	1a12aebb			Gratis par	###	Patio de L	###	Almería: A	< 100	arti	
38	136	Museo Ca	58562	Roscos	A-334, 04860 Olula de	66618: Pablo García	37.362626	f6836cc2			No dispon	###	Museo Ca	###	Olula del B	< 100	mus	
39	1371	Museo Ar	49130	Caracul	Cementerio, 11, 46300	62930: Lucrecia Baster	39.889130	5f5501dc			Gratis par	###	Casa de la	###	Olivia: V	< 100	pin	

# Selección de la variable de texto y var. suplementarias



Variables de texto analizadas:

- Autores de las obras
- Lugares de exposición: museos, galerías, salas de arte, etc.
- Comisarios de las exposiciones o curadores
- Coleccionistas
- Entidades organizadoras
- Entidades Financiadoras

# Ranking de artistas: listados alfabético y por ocurrencias

**Text Mining Analysis Results**

Corpus imported on NULL. Language: es.

Source: spreadsheet file C:/Users/user/Desktop/20160829\_EXPORTACIONES/20160829\_EXPORT\_PEPE\_VARIOS\_CRITERIOS\_SIN\_LIMITAR\_255.csv.

3391 documents and 10679 terms.

Processing options:

- Ignore case: enabled.
- Remove punctuation: disabled.
- Remove digits: disabled.
- Remove stopwords: disabled.
- Apply stemming: disabled.

**Terms dictionary sorted alphabetically**

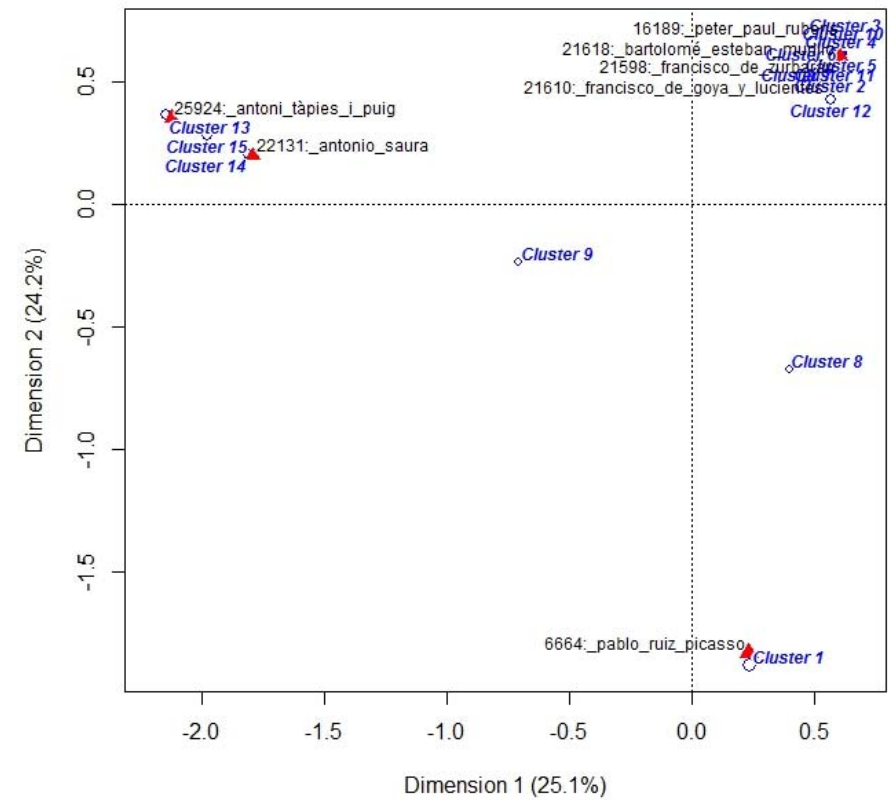
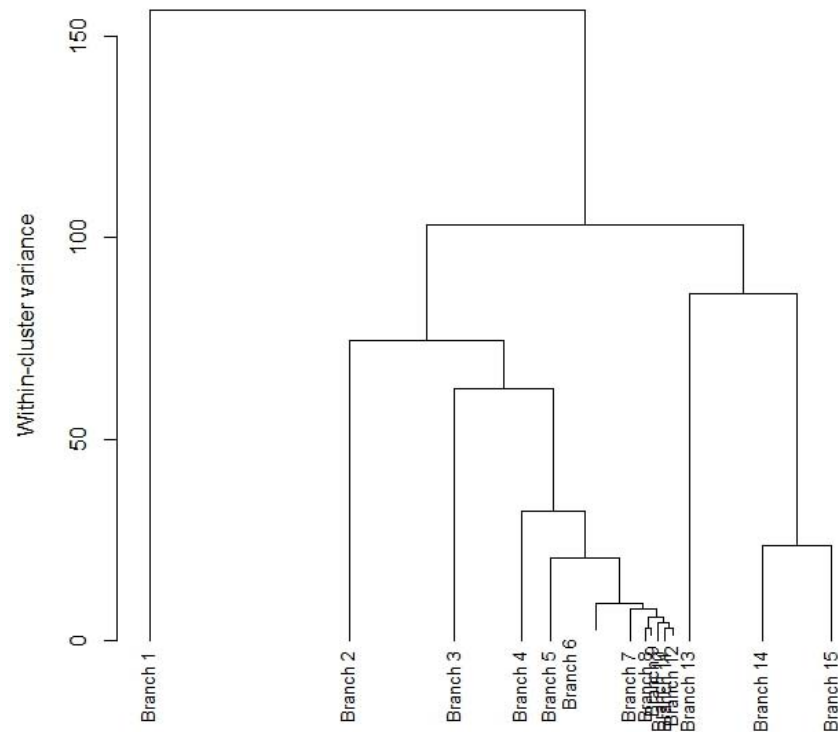
	Occurrences
16189_peter_paul_rubens	42
16195_claudio_de_lorena	18
16196_pieter_brueghel	4
16200_josé_garcía_ramos	1
19461_gonzalo_bilbao	3
21592_luis_de_morales	7
21598_francisco_de_zurbarán	35
21605_anton_van_dyck	14
21610_francisco_de_goya_y_lucientes	71
21614_julio_romero_de_torres	10
21618_bartolomé_esteban_murillo	36
21784_rafael_esteve_y_vilella	1
21785_luigi_pellegrino_scaramuccia	1
21786_paolo_veronese	11
21788_pascual_pedro_moles	1
21821_hermanos_penicaud	1
21823_juan_de_valdés_leal	10
21949_alberto_durero	17

**Terms dictionary sorted by number of occurrences**

	Occurrences
55293_[sin_determinar]	218
6664_pablo_ruiz_picasso	149
21610_francisco_de_goya_y_lucientes	71
23812_colectivo_sin_determinar	69
25924_antoni_tàpies_i_puig	60
7319_salvador_dalí	56
7677_joan_miró	55
16189_peter_paul_rubens	42
22131_antonio_saura	40
7675_diego_velázquez	40
25538_joaquín_sorolla_y_bastida	37
21618_bartolomé_esteban_murillo	36
21598_francisco_de_zurbarán	35
25463_equipo_crónica_o_crónicas_de_la_realidad	35
24401_ignasi_aballí	33
34097_josé_de_ribera_y_cucó	32
6527_el_greco	32
22112_luis_gordillo	31
25536_ignacio_pinazo_camarlench	31
24525_pablo_palazuelo_de_la_peña	30
27348_max_ernst	29
22103_eduardo_arroyo_rodríguez	27
22106_carmen_calvo	27
24436_antoni_muntadas	27
25223_eusebio_sempere	27
26582_robert_rauschenberg	27
27345_julio_gonzález	27
7264_tiziano	27
22113_cristina_iglesias_fernández_berrido	26
25477_juan_uslé	26
25854_txomin_badiola	26

# Clustering y Correspondence Analysis

Clusters dendrogram



# Estudio de la relación de asociación entre operadores de telefonía, tipos de red, potencia de la señal recibida y modelo de móvil, según el conjunto de datos públicos “Málaga CitySense - Mayo 2015”.

---

Análisis de Correspondencias, simple y múltiple, entre variables cualitativas nominales y cuantitativas, con R.TeMiS.

---

**José Pino-Díaz<sup>1</sup> y Adrián Pino-Martínez<sup>2</sup>**

1.- Universidad de Málaga, Andalucía Tech, Escuela de Ingenierías Industriales, Campus de Teatinos s/n, 29071 Málaga, España.

2.- Universidad de Granada. Escuela Técnica Superior de Ingeniería Informática y Telecomunicación.

# Conjunto de datos Málaga City Sense – Mayo 2015

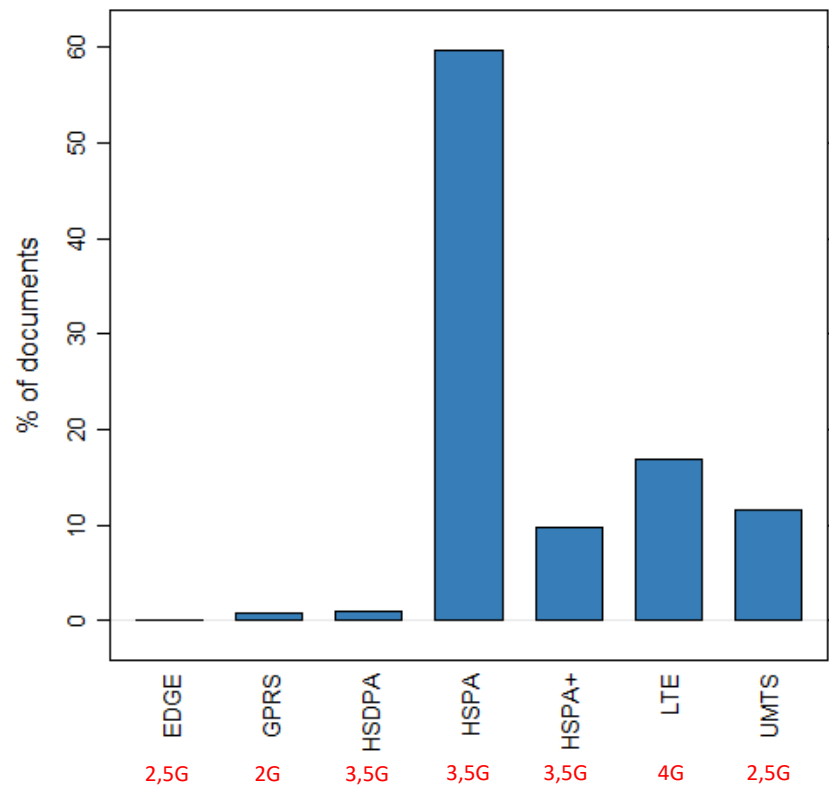
## (2.615 documentos)

user_id	sm_id	temperature	light	humidity	pressure	magnetism	speed	proximity	rotation	sound	latitude	longitude	altitude	date	acceleration	gravity	gsm_signal	network_type	operator	battery_level	battery_temp	device_model	app_version
779	224051	0	0	0	1011,0443	29,588032	0,75	0	0,07167665	41	36,7174054	-4,2258698	11	02/05/2015 22:23	0,950885	9,80665	85	LTE	Movistar	98	30	Nexus 5	5.1
779	224052	0	0	0	1011,0943	31,84587	0,75	0	0,20998499	12	36,7174054	-4,2258698	11	02/05/2015 22:24	0,9262662	9,80665	85	LTE	Movistar	98	30	Nexus 5	5.1
779	224554	0	4	0	1012,8803	91,28235	0	0	0,07524659	40	36,7165629	-4,2258665	113	05/05/2015 23:32	0,91972315	9,80665	-89	GPRS	Movistar	30	29	Nexus 5	5.1
779	224555	0	4	0	1012,8633	91,309525	0	0	0,18432996	27	36,7165629	-4,2258665	113	05/05/2015 23:32	0,9154848	9,80665	-89	GPRS	Movistar	30	29	Nexus 5	5.1
779	224556	0	0	0	1012,70337	88,25341	2,5495098	0	0,04029753	4	36,717197	-4,2258369	49	06/05/2015 0:25	0,989008	9,80665	85	LTE	Movistar	30	31	Nexus 5	5.1
779	224557	0	0	0	1012,65344	86,99213	2,5495098	0	0,04685998	0	36,717197	-4,2258369	49	06/05/2015 0:25	0,9750457	9,80665	85	LTE	Movistar	30	31	Nexus 5	5.1
779	224564	0	0	0	1012,66565	67,18326	0	0	0,03783182	11	36,7170811	-4,2258508	88	06/05/2015 3:25	1,0582652	9,80665	85	LTE	Movistar	100	27	Nexus 5	5.1
779	224572	0	0	0	1012,9033	65,54446	0	0	0,03855724	9	36,7170096	-4,2257206	189	06/05/2015 6:25	1,0585766	9,80665	85	LTE	Movistar	100	27	Nexus 5	5.1
779	224580	0	36	0	1014,1126	113,57103	0	0	0,15463388	10	36,7171022	-4,2259335	98	06/05/2015 9:25	0,91587377	9,80665	85	LTE	Movistar	100	26	Nexus 5	5.1
779	225066	0	0	0	1016,2115	70,03262	0	0	0,0320558	29	36,7170968	-4,2256066	63	07/05/2015 1:17	1,0071133	9,80665	85	LTE	Movistar	45	29	Nexus 5	5.1
779	225067	0	0	0	1016,13153	72,318146	0	0	0,03128334	5	36,7170968	-4,2256066	63	07/05/2015 1:17	1,0031581	9,80665	85	LTE	Movistar	45	29	Nexus 5	5.1
779	225070	0	0	0	1015,36194	675,21484	4,1231055	0	0,0345488	6	36,7170557	-4,2258745	100	07/05/2015 3:21	0,939506	9,80665	85	Desconocido		100	29	Nexus 5	5.1
779	225071	0	0	0	1015,32196	677,1207	4,1231055	0	0,03286815	0	36,7170557	-4,2258745	100	07/05/2015 3:22	0,94166875	9,80665	85	Desconocido		100	29	Nexus 5	5.1
779	225074	0	0	0	1015,4019	674,4052	4,1231055	0	0,03203495	0	36,7170557	-4,2258745	100	07/05/2015 6:21	0,95429116	9,80665	85	Desconocido		100	28	Nexus 5	5.1
779	225096	0	11	0	1012,8433	91,2772	0,25	0	0,15585934	34	36,7348617	-4,5574265	134	07/05/2015 10:48	0,9734392	9,80665	85	LTE	Movistar	79	38	Nexus 5	5.1
780	223991	0	0	0	0	47,062458	0,5	0	0	19	36,7182654	-4,4380165	69	01/05/2015 2:58	0,9814501	0	-79	HSPA	Vodafone_Es	26	27	LG-D320	4.4.2
780	223992	0	0	0	0	47,06445	0,5	0	0	1	36,7182654	-4,4380165	69	01/05/2015 2:58	0,941228	0	-79	HSPA	Vodafone_Es	26	27	LG-D320	4.4.2
780	223995	0	0	0	0	45,35003	0,5	0	0	2	36,7182654	-4,4380165	69	01/05/2015 5:58	0,9642124	0	-77	HSPA	Vodafone_Es	17	27	LG-D320	4.4.2
780	223996	0	0	0	0	45,714466	0,5	0	0	0	36,7182654	-4,4380165	69	01/05/2015 5:58	0,9623452	0	-77	HSPA	Vodafone_Es	17	27	LG-D320	4.4.2
780	223999	0	0	0	0	46,050243	0,5	0	0	2	36,7182654	-4,4380165	69	01/05/2015 8:58	0,97587115	0	-75	HSPA	Vodafone_Es	9	27	LG-D320	4.4.2
780	224000	0	0	0	0	46,69181	0,5	0	0	7	36,7182654	-4,4380165	69	01/05/2015 8:58	0,9758473	0	-75	HSPA	Vodafone_Es	9	28	LG-D320	4.4.2
780	224005	0	0	0	0	27,316433	0,5	0	0	12	36,7182654	-4,4380165	69	01/05/2015 13:59	0,9715405	0	-73	HSPA	Vodafone_Es	100	27	LG-D320	4.4.2
780	224007	0	0	0	0	18,836468	0,5	0	0	42	36,7182654	-4,4380165	69	01/05/2015 16:58	0,9430752	0	-69	HSPA	Vodafone_Es	97	29	LG-D320	4.4.2
780	224008	0	0	0	0	20,164946	0,5	0	0	19	36,7182654	-4,4380165	69	01/05/2015 16:59	0,9679614	0	-69	HSPA	Vodafone_Es	97	29	LG-D320	4.4.2
780	224009	0	0	0	0	20,629166	0,5	0	0	42	36,7182654	-4,4380165	69	01/05/2015 19:58	0,94694906	0	-73	HSPA	Vodafone_Es	90	29	LG-D320	4.4.2
780	224010	0	0	0	0	22,138767	0,5	0	0	18	36,7182654	-4,4380165	69	01/05/2015 19:59	0,9813003	0	-73	HSPA	Vodafone_Es	90	29	LG-D320	4.4.2
780	224016	0	0	0	0	16,902292	0,5	0	0	41	36,7182654	-4,4380165	69	01/05/2015 22:58	0,9758167	0	-75	HSPA	Vodafone_Es	82	29	LG-D320	4.4.2
780	224017	0	0	0	0	18,784967	0,5	0	0	18	36,7182654	-4,4380165	69	01/05/2015 22:59	0,94709975	0	-75	HSPA	Vodafone_Es	82	29	LG-D320	4.4.2
780	224023	0	0	0	0	20,665491	0,5	0	0	41	36,7182654	-4,4380165	69	02/05/2015 1:58	0,9680434	0	-69	HSPA	Vodafone_Es	74	29	LG-D320	4.4.2
780	224024	0	0	0	0	21,566757	0,5	0	0	18	36,7182654	-4,4380165	69	02/05/2015 1:59	0,95657164	0	-69	HSPA	Vodafone_Es	74	29	LG-D320	4.4.2
780	224025	0	0	0	0	21,177818	0,5	0	0	4	36,7182654	-4,4380165	69	02/05/2015 4:58	0,9740105	0	-69	HSPA	Vodafone_Es	67	29	LG-D320	4.4.2
780	224026	0	0	0	0	21,204363	0,5	0	0	0	36,7182654	-4,4380165	69	02/05/2015 4:59	0,94705206	0	-69	HSPA	Vodafone_Es	67	29	LG-D320	4.4.2
780	224029	0	0	0	0	19,978113	0,5	0	0	5	36,7182654	-4,4380165	69	02/05/2015 7:58	0,94330406	0	-71	HSPA	Vodafone_Es	59	28	LG-D320	4.4.2
780	224030	0	0	0	0	20,737947	0,5	0	0	0	36,7182654	-4,4380165	69	02/05/2015 7:59	0,94900703	0	-71	HSPA	Vodafone_Es	59	28	LG-D320	4.4.2
780	224035	0	0	0	0	20,322401	0,5	0	0	34	36,7182654	-4,4380165	69	02/05/2015 10:58	0,9815178	0	-71	HSPA	Vodafone_Es	51	28	LG-D320	4.4.2
780	224036	0	0	0	0	21,202888	0,5	0	0	12	36,7182654	-4,4380165	69	02/05/2015 10:59	0,94916344	0	-71	HSPA	Vodafone_Es	51	28	LG-D320	4.4.2
780	224037	0	0	0	0	22,2977	0,5	0	0	41	36,7182654	-4,4380165	69	02/05/2015 13:58	0,9454442	0	-71	HSPA	Vodafone_Es	43	28	LG-D320	4.4.2

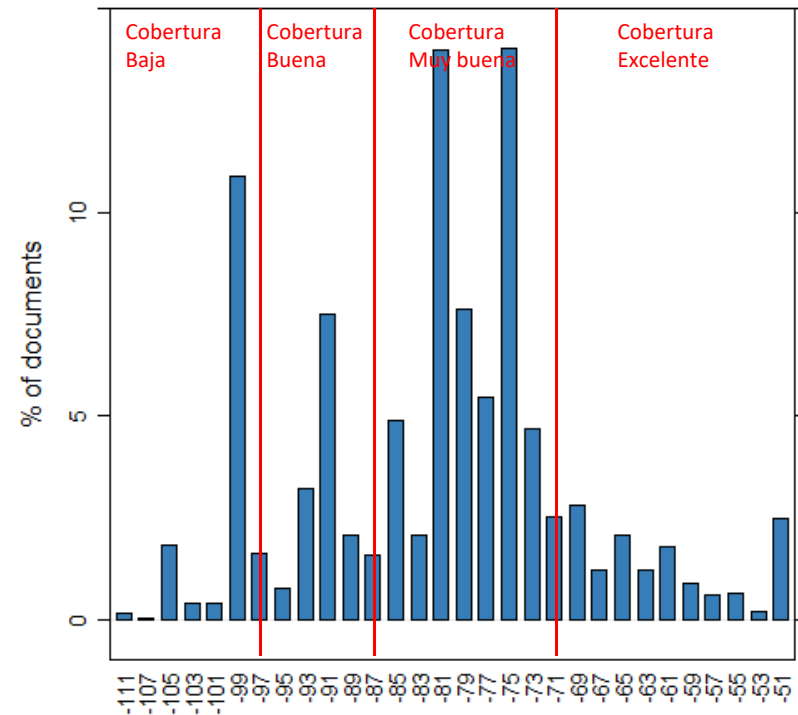


## Distribución por variables (*Network\_type* y *GSM\_signal*)

Distribution of data by network type



Distribution of data by GSM signal



*“Reliability and Failure\*”,*  
patentes del año 2018  
en la BD Espacenet

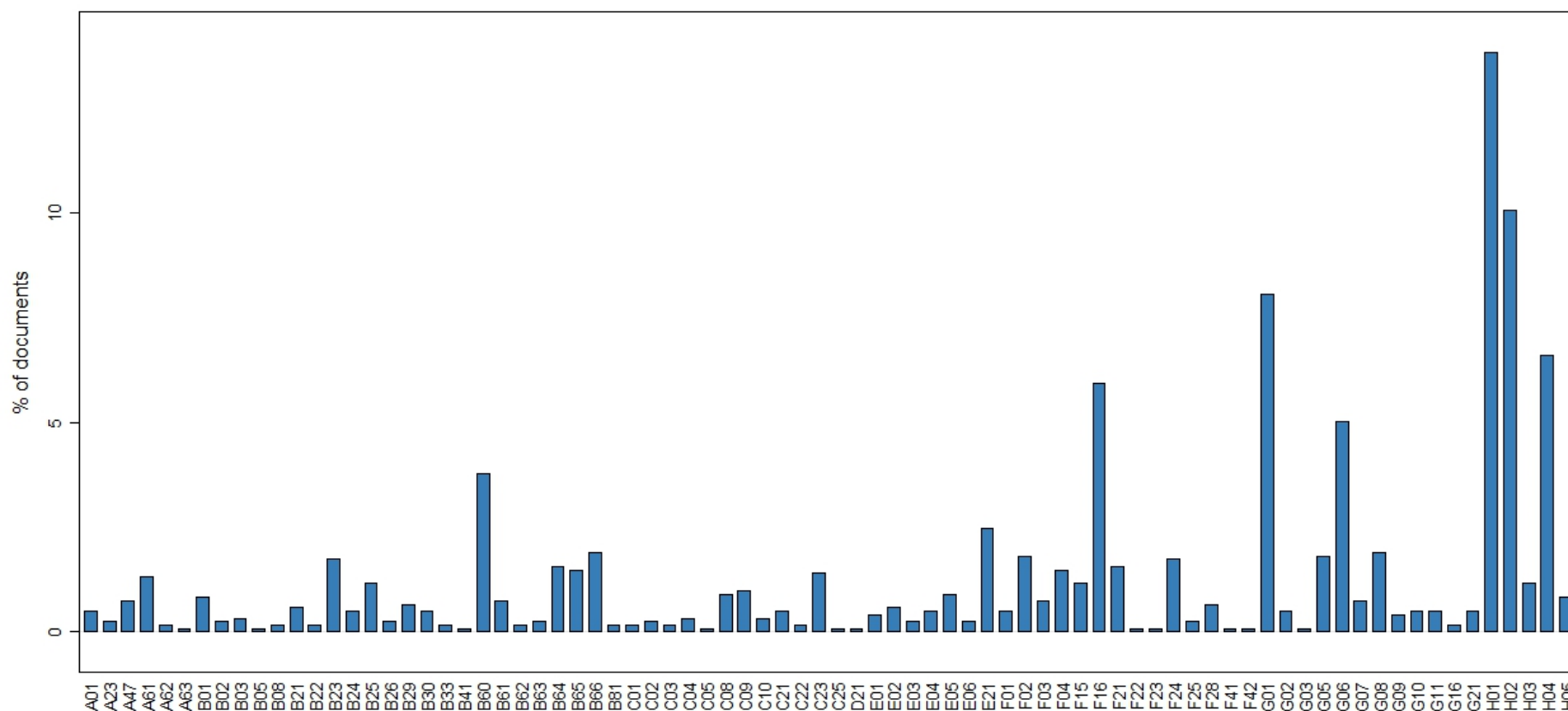
Clasificación y análisis factorial de  
correspondencias de los códigos CIT de las  
patentes

## Patentes sobre “*Reliability and Failure*” en Espacenet del año 2018<sup>1</sup>

Sección	Núm. patentes	%	Título de la sección de la CIP
A	37	3,05	NECESIDADES CORRIENTES DE LA VIDA
B	212	17,45	TECNICAS INDUSTRIALES DIVERSAS; TRANSPORTES
C	65	5,35	QUIMICA; METALURGIA
D	1	0,08	TEXTILES; PAPEL
E	65	5,35	CONSTRUCCIONES FIJAS
F	196	16,13	MECANICA; ILUMINACION; CALEFACCION; ARMAMENTO; VOLADURA
G	245	20,16	FÍSICA
H	394	32,43	ELECTRICIDAD
Total	1215	100,00	

# Patentes sobre “*Reliability and Failure*” en Espacenet del año 2018<sup>1</sup>

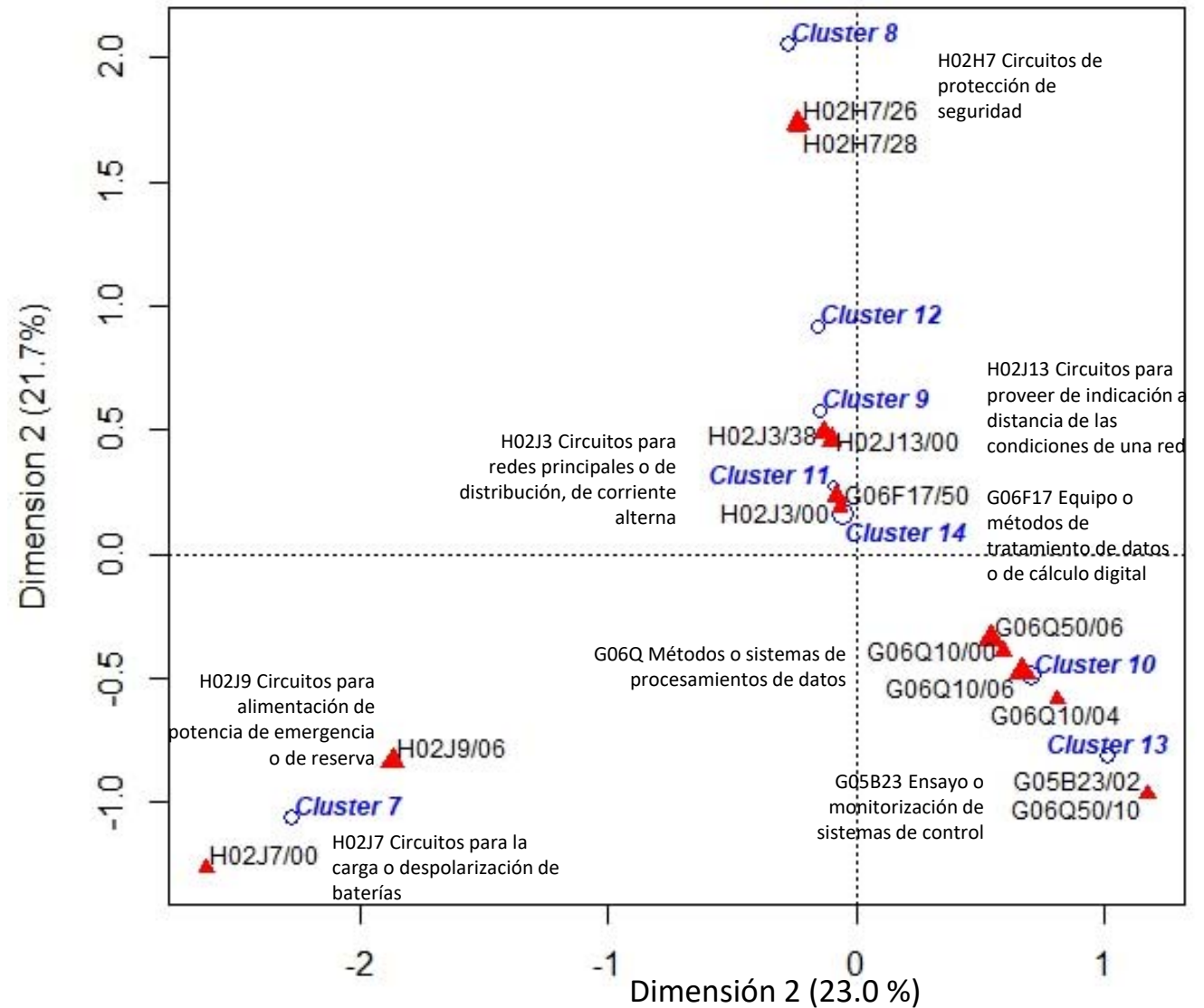
Distribution of CIP by Sección-Primer\_nivel



# Patentes sobre “*Reliability and Failure*” en Espacenet del año 2018<sup>1</sup>

Sección-1Nivel	Nº_Patentes	Categoría
H01	168	ELEMENTOS ELECTRICOS BASICOS
H02	122	PRODUCCION, CONVERSION O DISTRIBUCION DE LA ENERGIA ELECTRICA
G01	98	METROLOGIA; ENSAYOS
H04	80	TECNICA DE LAS COMUNICACIONES ELECTRICAS
F16	72	ELEMENTOS O CONJUNTOS DE TECNOLOGIA; MEDIDAS GENERALES PARA ASEGURAR EL BUEN FUNCIONAMIENTO DE LAS MAQUINAS O INSTALACIONES;
G06	61	COMPUTO; CALCULO; CONTEO
B60	46	VEHICULOS EN GENERAL
E21	30	PERFORACION DEL SUELO O DE LA ROCA; EXPLOTACION MINERA
B66	23	ELEVACION; LEVANTAMIENTO; REMOLCADO
G08	23	SEÑALIZACION
F02	22	MOTORES DE COMBUSTION; PLANTAS MOTRICES DE GASES CALIENTES O DE PRODUCTOS DE COMBUSTION
G05	22	CONTROL; REGULACION
B23	21	MAQUINAS-HERRAMIENTAS; TRABAJO DE METALES NO PREVISTO EN OTRO LUGAR
F24	21	CALEFACCION; HORNILLAS; VENTILACION
B64	19	AERONAVES; AVIACION; ASTRONAUTICA
F21	19	ILUMINACION
B65	18	TRANSPORTE; EMBALAJE; ALMACENADO; MANIPULACION DE MATERIALES DELGADOS O FILIFORMES
F04	18	MAQUINAS DE LIQUIDOS DE DESPLAZAMIENTO POSITIVO; BOMBAS PARA LIQUIDOS O PARA FLUIDOS COMPRESIBLES
C23	17	REVESTIMIENTO DE MATERIALES METALICOS; REVESTIMIENTO DE MATERIALES CON MATERIALES METALICOS; TRATAMIENTO QUIMICO DE LA SUPERFICIE
A61	16	CIENCIAS MEDICAS O VETERINARIAS; HIGIENE
B25	14	HERRAMIENTAS MANUALES; HERRAMIENTAS DE MOTOR PORTATILES; MANGOS PARA UTENSILIOS MANUALES; UTILLAJE DE TALLER; MANIPULADORES
F15	14	DISPOSITIVOS ACCIONADORES POR PRESION DE UN FLUIDO; HIDRAULICA O NEUMATICA EN GENERAL
H03	14	CIRCUITOS ELECTRONICOS BASICOS
C09	12	COLORANTES; PINTURAS; PULIMENTOS; RESINAS NATURALES; ADHESIVOS; COMPOSICIONES NO PREVISTAS EN OTRO LUGAR; APLICACIONES DE LOS MATERIALES
C08	11	COMPUESTOS MACROMOLECULARES ORGANICOS; SU PREPARACION O PRODUCCION QUIMICA; COMPOSICIONES BASADAS EN COMPUESTOS MACROMOLECULARES
E05	11	CERRADURAS; LLAVES; ACCESORIOS DE PUERTAS O VENTANAS; CAJAS FUERTES
B01	10	PROCEDIMIENTOS O APARATOS FISICOS O QUIMICOS EN GENERAL
H05	10	TECNICAS ELECTRICAS NO PREVISTAS EN OTRO LUGAR

Diagrama AC (zoom 3, eliminando los clústeres 3 y 5)





# Aplicación de minería de textos en análisis de las 215 respuestas de encuestas abiertas



Con el objeto de mejorar futuros MOOC y de continuar avanzando en una cultura de la innovación te pedimos valores de 1 a 10 los siguientes aspectos del que has realizado:

# Vamos a analizar las respuestas al apartado 1 de la encuesta: “Nivel de consecución de los siguientes Objetivos del MOOC”

Con el objeto de mejorar futuros MOOC y de continuar avanzando en una cultura de la innovación te pedimos valores de 1 a 10 los siguientes aspectos del que has realizado:

## **1. Nivel de consecución de los siguientes Objetivos del MOOC**

- Sensibilizar y motivar a los y las participantes para extender una cultura de la innovación en la Junta de Andalucía.
- Identificar las claves de la innovación en el sector público.
- Facilitar el intercambio, la interacción y el debate colectivo entre participantes en el MOOC, creando dinámicas de trabajo colaborativo.
- Familiarizarse con el uso de herramientas tecnológicas de información y comunicación
- Conocer experiencias de innovación en el sector público.
- Facilitar recursos y herramientas que permitan definir e implementar retos y proyectos de innovación.

Comentarios:

# Resultados

Comentarios al apartado 1 de la encuesta: "Nivel de consecución de los siguientes Objetivos del MOOC"			
Grupos	Términos significativos	A favor (+)	En contra (-)
Grupo 1	"experiencia"	Experiencia positiva. Me ha gustado la experiencia.	Excesiva burocracia.
Grupo 2	"curso"	Me ha entusiasmado. Fascinante. Me ha encantado. Curso motivador. Curso diferente. Curso completo.	Me ha decepcionado un poco la insignia. Me hubiera gustado tener un certificado del curso para poder aportarlo como mérito
Grupo 3	"herramientas"	Me ha aportado conocimientos, herramientas y recursos muy útiles. Me he iniciado en el uso de nuevas herramientas que desconocía.	Dificultad en el manejo de herramientas tecnológicas que facilitan el trabajo colaborativo
Grupo 4	"innovación"	Buena toma de contacto con el tema de la innovación. Se da a conocer la innovación en el trabajo del empleado público.	He encontrado a faltar experiencias de innovación en sistemas de gestión de RR.HH Parecía más un curso de redes sociales y autobombo que de innovación en sí
Grupo 5	"interesante"	Interesante por lo innovador en el método didáctico	
Grupo 6	"nuevas"	Abre nuevos horizontes. He hecho cosas nuevas. He practicado con nuevas tecnologías	
Grupo 7	"aprendido", "hacer"	He aprendido a hacer cosas nuevas. He conocido iniciativas muy interesantes de compañeros.	
Grupo 8	"innovar"	Se puede innovar en el trabajo. Innovar es importante.	Más empeño en twittear que en innovar.
Grupo 9	"puestos"		Impedimentos técnicos para realizar el curso en algunos puestos de trabajo. La realidad en los puestos de trabajo es otra a la planteada en el curso.
Grupo 10	"complicado"		Curso complicado para personas sin mínimos conocimientos en informática. La elevada participación y debate complica el resultado. En la Admon es complicado innovar desde la perspectiva del curso.
Grupo 11	"gran"	Gran seguimiento semanal Gran capacidad de motivación	
Grupo 12	"conseguido", "objetivo"	Objetivo conseguido	Decepcionada por la evaluación entre pares
Grupo 13	"objetivos"	Grado satisfactorio por conseguir los objetivos	
Grupo 14	"personas"	Curso muy motivador para personas con ganas de mejorar Curso muy activo donde personas novatas consiguen objetivos	No se expide certificado del curso para todas las personas
Grupo 15	"ideas"	Se comparten y se toman ideas.	

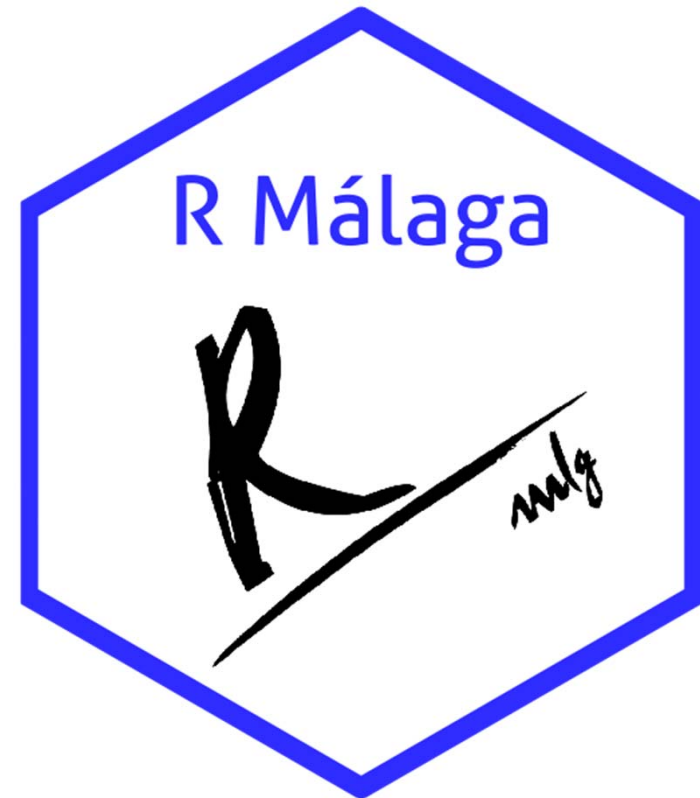
Toma de decisiones -> Para mejorar futuros MOOC y continuar avanzando en una cultura de la innovación en la Admon

La puntuación media de los seis objetivos del MOOC es de 8,3 sobre 10. Ahora bien, hay aspectos a mejorar. Según los comentarios al apartado 1 de la encuesta:

- a. Mejorar la insignia del curso.
- b. Expedir certificado para todos los participantes que lo superen, independientemente del tipo de Admon donde trabajen.
- c. Homogeneizar los grupos de participantes según el nivel de usuario en informática.
- d. Incluir y explicar experiencias de innovación en RR.HH.
- e. Facilitar la realización del curso, resolver los impedimentos técnicos en los centros.
- f. Acercar los planteamientos del curso a la realidad cotidiana de los centros.
- g. Optimizar el número de participantes para agilizar la participación y los debates.

¿Preguntas?...  
¡Comentarios!...

Muchas Gracias



**Prof. Dr. José Pino Díaz**

Universidad de Málaga

Andalucía Tech

Escuela de Ingenierías Industriales

Departamento de Economía y Administración de Empresas,  
Grupos de investigación Techné (UGR) e iArtHis-Lab (UMA)

Campus de Teatinos s/n, 29071 Málaga, España